Post Data Collection Processing Status

Presentation to the National Advisory Committee on Racial, Ethnic and Other Populations
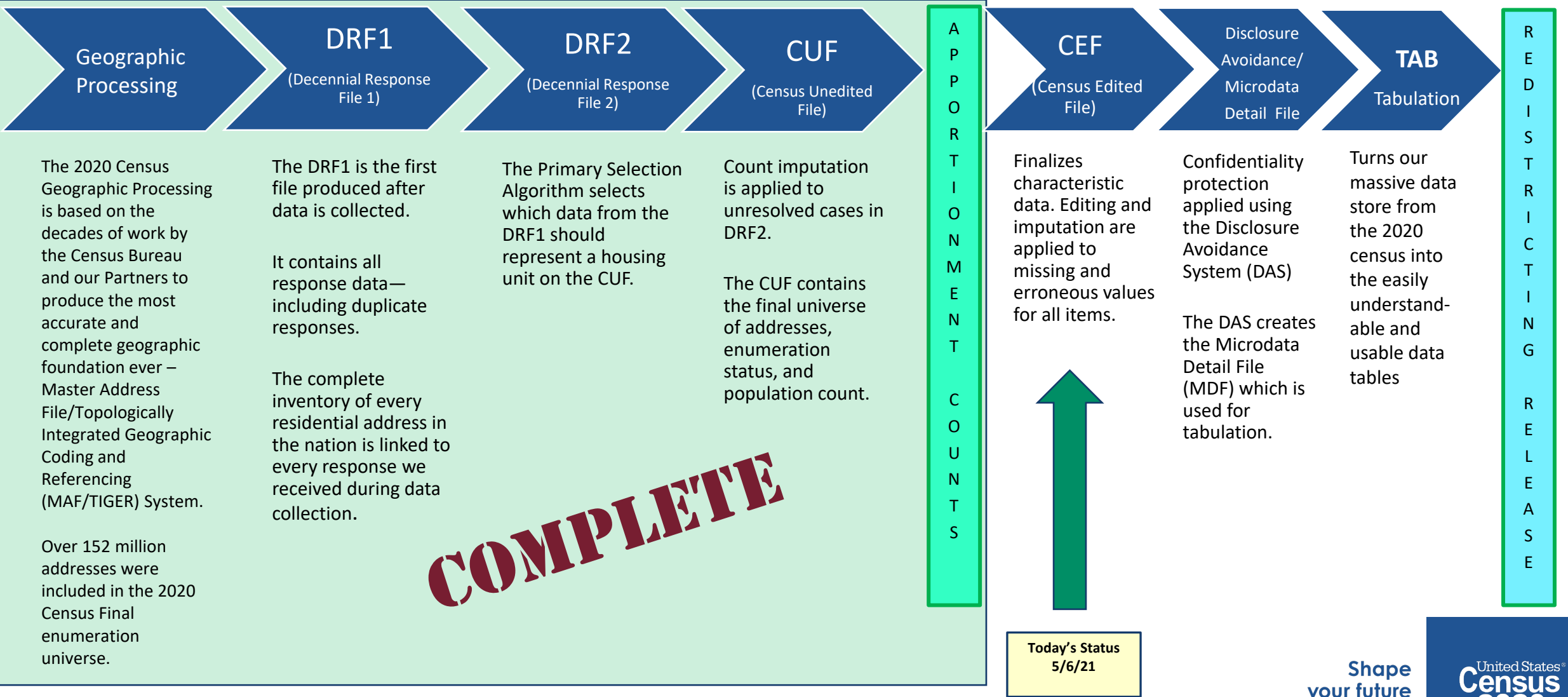
May 6, 2021

Michael Thieme, Assistant Director Decennial Census Programs, Systems and Contracts

Shape
your future
START HERE >

United States®
Census
2020

2020 Census
Post Data Collection Processing Flow

Geographic Processing

The 2020 Census Geographic Processing is based on the decades of work by the Census Bureau and our Partners to produce the most accurate and complete geographic foundation ever – Master Address File/Topologically Integrated Geographic Coding and Referencing (MAF/TIGER) System.

Over 152 million addresses were included in the 2020 Census Final enumeration universe.

DRF1 (Decennial Response File 1)

The DRF1 is the first file produced after data is collected.

It contains all response data—including duplicate responses.

The complete inventory of every residential address in the nation is linked to every response we received during data collection.

DRF2 (Decennial Response File 2)

The Primary Selection Algorithm selects which data from the DRF1 should represent a housing unit on the CUF.

CUF (Census Unedited File)

Count imputation is applied to unresolved cases in DRF2.

The CUF contains the final universe of addresses, enumeration status, and population count.

APPORTIONMENT COUNTS

COMPLETE

CEF (Census Edited File)

Finalizes characteristic data. Editing and imputation are applied to missing and erroneous values for all items.

Today's Status 5/6/21

Disclosure Avoidance/ Microdata Detail File

Confidentiality protection applied using the Disclosure Avoidance System (DAS)

The DAS creates the Microdata Detail File (MDF) which is used for tabulation.

TAB Tabulation

Turns our massive data store from the 2020 census into the easily understand-able and usable data tables

REDISTRICTING RELEASE

Shape your future START HERE >

United States® Census 2020

2020 Census
Apportionment and Redistricting
Key Milestones

Processing Step	Start	Finish
Decennial Response File 1 (DRF1)	10/29/2020	12/26/2020
Decennial Response File 2 (DRF2)	12/26/2020	2/24/2021
Census Unedited File (CUF)	2/25/2021	3/10/2021
Apportionment Preparation and Release*	3/12/2021	4/26/2021
Census Edited File (CEF)	4/20/2021	6/24/2021
Disclosure Avoidance Application/Microdata Detail File	6/25/2021	7/18/2021
Tabulation File Processing	7/19/2021	8/16/2021
Redistricting Preparation and Release	8/17/2021	9/30/2021

COMPLETE

Shape your future
START HERE >

United States®
Census
2020

2020 Census
Additional Details on Remaining Processing Steps

The Census Edited File (or CEF)

- Provides the much-more detailed data required for redistricting

- Complex processing for the race, ethnicity and age information that states must have to conduct redistricting

- Since CUF was primarily concerned with population counts, the detailed respondent information processed for the CEF phase can still be conflicting or contradictory, so it needs further attention in this phase

- During CEF processing, missing characteristic data are accounted for using a statistical process called *characteristic* imputation

- Iterative run and review cycles to ensure processing is working correctly

Shape
your future
START HERE >

United States®
Census
2020

2020 Census
Additional Details on Remaining Processing Steps

The Microdata Detail File (or MDF)

- The Census Bureau is required by law to protect the confidentiality of the information provided by respondents

- Processing for the MDF revolves around keeping your data private and protected

- For the first time, the Census Bureau is planning to use a statistical method called differential privacy to protect the confidentiality of individuals in the 2020 Census

- This stage is designed for and requires the entire nation be processed as whole

- Includes time for review and possible rerun to ensure the system is functioning as designed and your data are protected

Shape
your future
START HERE >

United States®
Census
2020

2020 Census
Additional Details on Remaining Processing Steps

Tabulation

- Turns our massive data store from the 2020 census into the easily understandable and usable data tables

- Makes census data easily accessible

- Accurately integrates geography, population, and characteristics in a huge number of combinations

- Requires a thorough review and validation to ensure the accuracy of these products

- During tabulation we add new data fields that make future data use and deeper understanding easier for data users

- Generate the specially tabulated datasets in the particular required formats that make these files easily usable by state redistricting people

Shape your future START HERE >

United States® Census 2020

2020 Census
Additional Details on Remaining Processing Steps

Redistricting Release

- Review, prepare, load, and deliver the official redistricting data

- Perform final testing to ensure that the software and the data conversion have all worked correctly

- Create DVDs and Flash Drives with each state's master image to include both the data and the end-user software that makes the data easy to use

- load "data.census.gov" Data Explorer tool with the entire nation's data and metadata that enables users to pull and display data easily on the web

- Deliver the physical materials via overnight mail to all states

- Make web-tool and data available to the official recipients and the public

Shape
your future
START HERE >

United States®
Census
2020

Questions